

Improved Update Operators for Lifting-Based Motion-Compensated Temporal Filtering

Christophe Tillier, Béatrice Pesquet-Popescu, *Member, IEEE*, and Mihaela van der Schaar, *Member, IEEE*

Abstract—Motion-compensated temporal filtering (MCTF) is an essential ingredient of the recently developed wavelet-based scalable video coding schemes. Lifting implementation of these decompositions represents a versatile tool for spatiotemporal optimizations, and several improvements have already been proposed in this framework. In this paper, we propose new spatiotemporal update operators based either on weighted averages of multiple connected pixels or on nonlinear filtering of these pixels. We demonstrate the improved performance of the proposed operators, both theoretically and experimentally.

Index Terms—Motion-compensated temporal filtering, nonlinear operators, spatiotemporal update operator, temporal lifting.

I. INTRODUCTION

SCALABLE video coding based on motion-compensated (MC) spatiotemporal ($t + 2D$) wavelet decompositions is becoming increasingly popular, as it provides coding performance competitive with state-of-the-art codecs, while trying to accommodate varying network bandwidths and different receiver capabilities (frame-rate, display size, CPU, etc.) and to provide solutions for network congestion or video server design. Block-based motion compensation (with fixed or variable block sizes) is used in most of the proposed schemes in this framework, due to its simplicity and the long experience that was developed around this tool. However, this technique raises, in the case of subband temporal filtering, some problems related to the nonsymmetry of the prediction relation: The motion vector field (MVF) estimated from a frame A to a frame B is not exactly the opposite of the MVF estimated from B to A . This leads to the so-called “unconnected” and “multiple connected” pixels [1], [2], which could produce annoying coding artefacts. They represent, respectively, the pixels in the reference frame not used for temporal prediction and those used to predict more than one pixel in the current frame. The use of mesh models for motion compensation [3] could be a solution to avoid this problem. However, the popularity of block-based motion estimation techniques plead in favor of finding a satisfactory way to process these pixels in order to provide high-efficiency video codecs. The case of unconnected pixels was considered,

Manuscript received May 8, 2004; revised August 20, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dipti Prasad Mukherjee.

C. Tillier and B. Pesquet-Popescu are with the Signal and Image Processing Department, Ecole Nationale Supérieure des Télécommunications and CNRS UMR 5141, Paris, France (e-mail: tillier,pesquet@tsi.enst.fr).

M. van der Schaar is with the Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA 95616 USA (e-mail: mvanderschaar@ece.ucdavis.edu).

Digital Object Identifier 10.1109/LSP.2004.840877

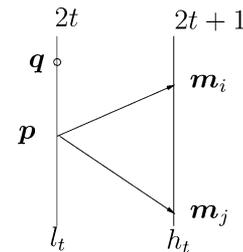


Fig. 1. Motion-compensated prediction in the two-band Haar scheme.

and a solution by lowpass transition spatial filtering was recently proposed in [4] and [5]. By performing the temporal multiresolution analysis in a nonlinear lifting framework [6], we have shown that it is possible to apply any transformation to the set of coefficients connected with a single pixel in the reference frame while preserving the perfect reconstruction of the scheme. In this paper, we are considering different methods for efficient processing of multiple connected pixels. We start with an analysis of the reconstruction error of the different types of pixels in an MC spatiotemporal two-band scheme. We prove that, rather than choosing a single pixel for performing the update, averaging all the connected pixels leads to a decrease of the reconstruction error globally and also individually on most of the pixels involved in the spatiotemporal filtering. The cases of normalized mean and unconstrained linear combination of values are analyzed. We confirm these theoretical results by simulations both on two-band and three-band motion-compensated temporal filtering (MCTF) codecs. Moreover, we compare this linear update with several nonlinear operators. Note that an independent work [7] proposes a global error analysis in the linear dyadic Haar lifting scheme, leading to a solution similar to what we called “the unconstrained average.”

The paper is organized as follows. In Section II, we introduce the proposed linear operators and analyze the resulting improvements in terms of reconstruction error compared with the existing scheme. In Section III, we analyze the triadic scheme, and in Section IV, we introduce the nonlinear operators for the update. Section V illustrates, via simulation results on both two-band and three-band systems, the coding gain of the proposed method. We conclude in Section VI.

II. RECONSTRUCTION ERROR AND AVERAGE UPDATE OPERATORS

We consider the two-band lifting-based MCTF framework, as described in [6]. As illustrated in Fig. 1, several pixels in frame x_{2t+1} , denoted by $(\mathbf{m}_i)_{i \in \{1, \dots, N\}}$, $N \geq 1$, can be predicted by motion compensation from the same pixel \mathbf{p} in the

frame x_{2t} . We say that \mathbf{p} is connected with each one of the pixels $(\mathbf{m}_i)_{i \in \{1, \dots, N\}}$. Let us also denote by \mathbf{q} an unconnected pixel in x_{2t} (not used for the prediction of any pixel in x_{2t+1}). The coefficients in the temporal detail frame h_t are obtained by

$$h_t(\mathbf{m}_i) = \frac{1}{\sqrt{2}}(\mathbf{m}_i - \mathbf{p}), \quad \forall i \in \{1, \dots, N\}.$$

Owing to the lifting formulation, we can involve all the connected pixels $(\mathbf{m}_i)_{i \in \{1, \dots, N\}}$ in the update, through a linear or nonlinear operator, while preserving the perfect reconstruction of the scheme. We propose here to construct the approximation subband coefficients by a weighted sum of the details in these points

$$l_t(\mathbf{q}) = \sqrt{2}\mathbf{q}, \quad l_t(\mathbf{p}) = \sqrt{2}\mathbf{p} + \sum_{i=1}^N \alpha_i h_t(\mathbf{m}_i)$$

where $(\alpha_i)_{i \in \{1, \dots, N\}}$ are some real constants. The synthesis equations are easily obtained:

$$\begin{aligned} \mathbf{m}_i &= \sqrt{2}h_t(\mathbf{m}_i) + \mathbf{p}, & \mathbf{q} &= \frac{1}{\sqrt{2}}l_t(\mathbf{q}) \\ \mathbf{p} &= \frac{1}{\sqrt{2}}\left(l_t(\mathbf{p}) - \sum_{i=1}^N \alpha_i h_t(\mathbf{m}_i)\right). \end{aligned}$$

In order to proceed with the reconstruction error analysis, we suppose that all the coefficients in the detail h_t frames are quantized with the same quantization step and are identically distributed. Let us denote by ε_{m_i} , ε_p , and ε_q the quantization errors, respectively, on $h_t(\mathbf{m}_i)$, $l_t(\mathbf{p})$, and $l_t(\mathbf{q})$, which will be considered statistically uncorrelated. The variances of these errors are given for all $i \in \{1, \dots, N\}$, by $\sigma_{\varepsilon_{m_i}}^2 = \sigma_{\varepsilon}^2$, $\sigma_{\varepsilon_p}^2$, and $\sigma_{\varepsilon_q}^2$. Then, from the synthesis equations, the variance of the reconstruction errors for the pixel \mathbf{p} will be related to the variance of the quantization errors as follows:

$$\sigma_p^2 = \frac{1}{2}\left(\sigma_{\varepsilon_p}^2 + \sum_{i=1}^N \alpha_i^2 \sigma_{\varepsilon}^2\right) \quad (1)$$

while that of any connected pixels \mathbf{m}_{i_0} is

$$\sigma_{\mathbf{m}_{i_0}}^2 = \frac{1}{2}\sigma_{\varepsilon_p}^2 + 2(1 - \alpha_{i_0})\sigma_{\varepsilon}^2 + \frac{1}{2}\sum_{i=1}^N \alpha_i^2 \sigma_{\varepsilon}^2.$$

The sum of reconstruction errors over all connected pixels is then

$$\sum_{i=1}^N \sigma_{\mathbf{m}_i}^2 = \left(\frac{N}{2}\sum_{i=1}^N \alpha_i^2 + 2N - 2\sum_{i=1}^N \alpha_i\right)\sigma_{\varepsilon}^2 + \frac{N}{2}\sigma_{\varepsilon_p}^2. \quad (2)$$

Note that the reconstruction error on the unconnected pixel \mathbf{q} does not depend on α_i , so it will not change for the different strategies discussed in the sequel.

A. Classical Update

In the previously introduced strategies, only one pixel, which we shall denote by \mathbf{m}_1 , is usually chosen for lowpass filtering (the choice often being done following the scanning order in the frame x_{2t+1} [2] or according to any other optimized criterion, as those exposed in [6]). This reduces to the particular case $\alpha_1 = 1$, $\alpha_j = 0$ for all $j > 1$. The reconstruction errors are then

$$\sigma_p^2 = \frac{1}{2}\left(\sigma_{\varepsilon_p}^2 + \sigma_{\varepsilon}^2\right), \quad \sigma_q^2 = \frac{1}{2}\sigma_{\varepsilon_q}^2 \quad (3)$$

$$\begin{aligned} \sigma_{m_1}^2 &= \frac{1}{2}\left(\sigma_{\varepsilon_p}^2 + \sigma_{\varepsilon}^2\right), & \sigma_{m_i}^2 &= \frac{1}{2}\left(\sigma_{\varepsilon_p}^2 + 5\sigma_{\varepsilon}^2\right) \\ &\forall i \neq 1 \end{aligned} \quad (4)$$

and the sum of the reconstruction errors of all considered pixels, except \mathbf{q} , is

$$\sigma_t^2 = \frac{N+1}{2}\sigma_{\varepsilon_p}^2 + \frac{5N-3}{2}\sigma_{\varepsilon}^2. \quad (5)$$

Recall from (4) that the reconstruction error of two pixels in the frame x_{2t+1} is different, depending on whether or not it has been used in the update operation. In the sequel, we describe an averaging method that leads to equal reconstruction errors for all pixels as well as to a decreased global reconstruction error.

B. Normalized Average of the Connected Pixels

If we set the condition $\sum_{i=1}^N \alpha_i = 1$ (this means that we really update with an average of all connected pixels and do not change the mean value of the data), then from (2), minimizing the mean error on these pixels amounts to minimizing the expression $(1/2)\sum_{i=1}^N \alpha_i^2 + 2 - (2/N)$. In the meantime, this also minimizes the reconstruction error on \mathbf{p} in (1). It is easy to see that the weights minimizing this expression under the unit sum constraint are $\alpha_i = 1/N$. The error compared with the previous case is, for all $N \geq 1$, as follows:

- larger on pixel \mathbf{m}_1 (the only one used in the previous scenario for updating). Indeed

$$\sigma_{m_1}^2 = \frac{1}{2}\sigma_{\varepsilon_p}^2 + \left(2 - \frac{3}{2N}\right)\sigma_{\varepsilon}^2 \geq \frac{1}{2}\left(\sigma_{\varepsilon_p}^2 + \sigma_{\varepsilon}^2\right)$$

- smaller on all the other pixels \mathbf{m}_i , $i > 1$, connected to \mathbf{p}

$$\sigma_{m_i}^2 = \frac{1}{2}\sigma_{\varepsilon_p}^2 + \left(2 - \frac{3}{2N}\right)\sigma_{\varepsilon}^2 \leq \frac{1}{2}\sigma_{\varepsilon_p}^2 + \frac{5}{2}\sigma_{\varepsilon}^2, \quad i > 1$$

- smaller on \mathbf{p}

$$\sigma_p^2 = \frac{1}{2}\sigma_{\varepsilon_p}^2 + \frac{1}{2N}\sigma_{\varepsilon}^2 \leq \frac{1}{2}\left(\sigma_{\varepsilon_p}^2 + \sigma_{\varepsilon}^2\right).$$

The sum of reconstruction errors of all N pixels connected to \mathbf{p} and that of \mathbf{p} is globally reduced. In this case, the total error becomes

$$\sigma_t^2 = \frac{N+1}{2}\sigma_{\varepsilon_p}^2 + \left(2N + \frac{1}{2N} - \frac{3}{2}\right)\sigma_{\varepsilon}^2$$

which is obviously lower than the variance error σ_t^2 , given by (5) in the classical case.

C. Unconstrained Average Update Operator

If we do not constrain the weights to sum up to 1, then an improved performance can be expected by minimizing the total reconstruction error

$$\left(\frac{N+1}{2}\sum_{i=1}^N \alpha_i^2 + 2N - 2\sum_{i=1}^N \alpha_i\right)\sigma_{\varepsilon}^2 + \frac{N+1}{2}\sigma_{\varepsilon_p}^2. \quad (6)$$

By deriving this expression w.r.t. each α_i , we get

$$\alpha_i = \frac{2}{N+1}, \quad \text{for all } i \in \{1, \dots, N\}. \quad (7)$$

Note that a different approach to introduce such an optimal update was independently proposed in [7].

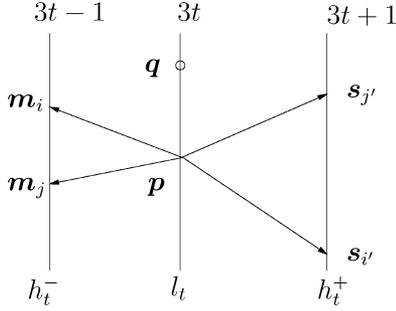


Fig. 2. Motion-compensated temporal prediction in a three-band lifting scheme.

In general, the error on a connected pixel is

$$\sigma_{m_i}^2 = \frac{1}{2}\sigma_{\varepsilon_p}^2 + 2 \left[1 - \frac{N+2}{(N+1)^2} \right] \sigma_{\varepsilon}^2. \quad (8)$$

As we have

$$\frac{1}{2} \leq 2 - \frac{2(N+2)}{(N+1)^2} \leq \frac{5}{2}, \quad \forall N \geq 1$$

we deduce that the pixel m_1 has a larger reconstruction error than in the original case, while all the other pixels connected with p , m_i , $i > 1$ have smaller reconstruction errors. In the same way, the pixel p also has a smaller error than before:

$$\sigma_p^2 = \frac{1}{2}\sigma_{\varepsilon_p}^2 + \frac{2N}{(N+1)^2}\sigma_{\varepsilon}^2 \leq \frac{1}{2}(\sigma_{\varepsilon_p}^2 + \sigma_{\varepsilon}^2), \quad \forall N \geq 1.$$

Globally, the reconstruction error has been minimized by the constraint we imposed since the sum of the variances is

$$\sigma_{i'}^2 = \frac{N+1}{2}\sigma_{\varepsilon_p}^2 + 2N \left(1 - \frac{1}{N+1} \right) \sigma_{\varepsilon}^2 \leq \sigma_{i'}^2 \leq \sigma_t^2.$$

III. EXTENSION OF THE PROPOSED APPROACH TO THREE-BAND STRUCTURES

A very similar analysis can be applied to nondyadic lifting schemes involving unidirectional MC temporal prediction, for example, the three-band ‘‘Haar-like’’ structure [8], whose temporal filtering structure is depicted in Fig. 2. The difference with the previous two-band case lies in the fact that there are two temporal detail subbands (h_t^+ , h_t^-), and a pixel p in frame $3t$ can be used for the prediction of multiple pixels, both in frame $3t - 1$ (like m_i , m_j), frame $3t + 1$ (like $s_{i'}$, and $s_{j'}$). Two solutions can be envisioned in this context. The first solution consists of computing the update by averaging the connected pixels in each detail subband independently and then taking the mean of the two results. The second possibility is to average all the connected pixels from the two frames with the same weights. A theoretical analysis similar to the one presented for the two-band scheme would plead in favor of the latter solution.

IV. NONLINEAR OPERATORS

In the update step, it is possible to apply to the N pixels $(m_i)_{i \in \{1, \dots, N\}}$ in frame x_{2t+1} connected to p a nonlinear function without affecting the perfect reconstruction property of the spatiotemporal lifting scheme. Among the plethora of possible nonlinear functions, two were more appealing:

- the median filter, for its well-known properties of eliminating outliers. Note that in the situation that is statistically the most likely, that of two connections, the median is equivalent to the average of the two values. Therefore, we do not expect a big difference in performance between these two operators.
- an adaptive weighted average of the connected pixels, with the weight being inversely proportional to the norm of the corresponding motion vector (MV). The weights in this case are

$$\alpha_i = \begin{cases} \frac{\|\mathbf{v}_{m_i}\|^{-1}}{\sum_{j=1}^N \|\mathbf{v}_{m_j}\|^{-1}}, & \text{if } \forall j, \|\mathbf{v}_{m_j}\| \neq 0 \\ 1, & \text{if } \|\mathbf{v}_{m_i}\| = 0 \\ 0, & \text{if } \exists j \neq i, \|\mathbf{v}_{m_j}\| = 0. \end{cases}$$

This solution makes the implicit hypothesis that the pixels having a small displacement w.r.t. the pixel p are more likely to correspond to the real displacement of the objects in the scene and, thus, to lead to approximation frames more meaningful and easier to encode. In particular, if one of the connected pixels in the frame x_{2t+1} has the same position as the pixel p , then it will be selected for the update operation, and all the others will be neglected.

V. EXPERIMENTAL RESULTS

We present simulation results for both the two-band and three-band schemes on three popular CIF sequences (‘‘stefan,’’ ‘‘foreman,’’ and ‘‘mobile’’) at 30 fps, selected for their different motion characteristics. The video codec used in the experiments consists in the MCTF (two-band or three-band), followed by a spatial decomposition of the temporal subbands with a 9/7 biorthogonal multiresolution analysis. The spatiotemporal wavelet coefficients are encoded with the MC-EZBC software [9]. This provides an embedded bitstream at full bitrate, which can be further cut at the required bitrates, resulting in the same quality as a direct encoding at the target bitrate. The same software is used to perform the motion estimation by hierarchical variable size block matching, with 1/8th pel accuracy and motion vector arithmetic encoding.

We compare in Table I several updating strategies. In the two-band MC filterbank these are as follows: first pixel used (‘‘First’’), the normalized mean (‘‘Mean’’), the nonnormalized mean (‘‘NN Mean’’), the median (‘‘Median’’), and the average weighted by taking into account the MV norm (‘‘MV Mean’’). In addition, for the three-band filterbank, we also compared with the average of the two values computed with connected pixels on each side (‘‘MMean’’) and the global mean of all connected pixels (‘‘BMean’’).

A first remark is that the normalized mean always brings an improvement, compared with the strategy of selecting the first pixel in the scanning order for the update operation. This improvement is up to 0.52 dB on the ‘‘stefan’’ sequence and two-band scheme and 0.37 dB on the ‘‘foreman’’ sequence, but it is less impressive on the ‘‘mobile’’ sequence, which has a more uniform motion. This gives rise to less multiple connected

TABLE I
RD (RATE IN KILOBITSECONDS, PSNR IN DECIBELS) COMPARISON BETWEEN SEVERAL UPDATING STRATEGIES, FOR TWO-BAND AND THREE-BAND MC FILTERBANKS AND THREE TEST SEQUENCES

| “stefan” 2B | 400kbs | 600kbs | 800kbs | 1000kbs | 1500kbs |
|--------------|--------|--------|--------|---------|---------|
| First | 23.56 | 26.89 | 28.67 | 30.03 | 32.73 |
| Mean | 23.82 | 27.35 | 29.13 | 30.55 | 33.22 |
| NN Mean | 23.84 | 27.31 | 29.07 | 30.49 | 33.17 |
| MV Mean | 23.75 | 27.19 | 28.98 | 30.39 | 33.08 |
| Median | 23.80 | 27.30 | 29.08 | 30.49 | 33.15 |
| “stefan” 3B | 400kbs | 600kbs | 800kbs | 1000kbs | 1500kbs |
| First | 24.82 | 27.19 | 28.98 | 30.39 | 33.08 |
| MMean | 24.90 | 27.91 | 29.61 | 31.01 | 33.55 |
| BMean | 24.92 | 27.92 | 29.63 | 31.03 | 33.57 |
| MV Mean | 24.88 | 27.86 | 29.55 | 30.96 | 33.50 |
| Median | 24.90 | 27.89 | 29.59 | 31.00 | 33.53 |
| “foreman” 2B | 200kbs | 400kbs | 600kbs | 800kbs | 1500kbs |
| First | 29.33 | 33.40 | 35.33 | 36.62 | 39.66 |
| Mean | 29.41 | 33.70 | 35.70 | 36.99 | 39.99 |
| NN Mean | 29.38 | 33.67 | 35.60 | 36.94 | 39.94 |
| MV Mean | 29.33 | 33.60 | 35.53 | 36.87 | 39.87 |
| Median | 29.40 | 33.67 | 35.61 | 36.95 | 39.93 |
| “foreman” 3B | 200kbs | 400kbs | 600kbs | 800kbs | 1500kbs |
| First | 30.34 | 34.29 | 35.98 | 37.33 | 40.23 |
| MMean | 30.40 | 34.40 | 36.09 | 37.43 | 40.32 |
| BMean | 30.42 | 34.41 | 36.11 | 37.44 | 40.32 |
| MV Mean | 30.37 | 34.36 | 36.06 | 37.40 | 40.29 |
| Median | 30.39 | 34.39 | 36.08 | 37.42 | 40.31 |
| “mobile” 2B | 200kbs | 400kbs | 600kbs | 800kbs | 1500kbs |
| First | 18.59 | 26.18 | 28.99 | 30.76 | 34.23 |
| Mean | 18.67 | 26.31 | 29.17 | 30.91 | 34.37 |
| NN Mean | 18.65 | 26.30 | 29.15 | 30.89 | 34.35 |
| MV Mean | 18.63 | 26.24 | 29.10 | 30.85 | 34.33 |
| Median | 18.67 | 26.30 | 29.16 | 30.90 | 34.36 |
| “mobile” 3B | 200kbs | 400kbs | 600kbs | 800kbs | 1500kbs |
| First | 22.26 | 28.48 | 30.64 | 32.12 | 35.10 |
| MMean | 22.29 | 28.53 | 30.68 | 32.16 | 35.14 |
| BMean | 22.28 | 28.54 | 30.69 | 32.16 | 35.14 |
| MV Mean | 22.26 | 28.50 | 30.66 | 32.15 | 35.13 |
| Median | 22.29 | 28.53 | 30.68 | 32.16 | 35.14 |

pixels and, thus, reduces the impact of the proposed update strategy. Indeed, the new update operators act only on multiple connected pixels, which represent globally a small percentage of the pixels involved in the temporal decomposition. For example, on “stefan,” this concerns 15.8% pixels, on “foreman,” 11.1%, and on “mobile,” 3.6% only.

For the three-band scheme, the peak signal-to-noise ratio (PSNR) differences between the different update strategies is up to 0.62 dB on the “stefan” sequence but generally much smaller than for the two-band decomposition since in the three-band scheme, the bidirectional update acts, from the beginning, as an average operator. In this case, however, one can note a slight improvement when using the global mean.

A second remark is that the difference between the normalized and the non-normalized average is very small in all situations but better for the normalized mean. This may be explained by the fact that in our theoretical analysis, we considered a fixed variance for the quantization noise. However, in

simulations, we have a fixed bitrate, and according to the quality of each representation, the quantization step varies for the different approaches. The comparison of the median operator with the linear averaging is slightly in favor of the latter one. The average, taking into account the norm of the corresponding motion vectors, has a slightly worse performance. This strategy allows for small displacements, which may lead to a larger penalty on sequences with complex motion, like “stefan” and “foreman.” Note, however, that all the proposed improved operators outperform the classical update procedure.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have introduced new spatiotemporal update operators for lifting-based motion-compensated temporal filtering. For the first one, performing a linear averaging of the multiple connected pixels, we showed both theoretically and by simulations that it provides improved coding performance for a variety of sequences. We have compared it with two other non-linear operators and highlighted its advantages, especially for high-motion sequences.

Extensions of the proposed spatiotemporal processing schemes can be employed for improving the resiliency of wavelet video compression schemes, depending on the content characteristics. Also, various averaging techniques, in combination with a multihypothesis approach, can be employed across spatiotemporal resolutions for improved scalability.

ACKNOWLEDGMENT

The authors would like to thank the reviewers, the Associate Editor and the Editor-in-Chief, for their valuable comments. M. van der Schaar would also like to acknowledge her discussions with Prof. B. Girod on wavelet video coding improvements.

REFERENCES

- [1] J.-R. Ohm, “Three-dimensional subband coding with motion compensation,” *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 559–589, Sep. 1994.
- [2] S. Choi and J. Woods, “Motion-compensated 3-D subband coding of video,” *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 155–167, Feb. 1999.
- [3] A. Secker and D. Taubman, “Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation,” in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2002.
- [4] K. Hanke, J.-R. Ohm, and T. Ruster, “Adaptation of filters and quantization in spatiotemporal wavelet coding with motion compensation,” in *Picture Coding Symp.*, St. Malo, France, Apr. 2003, pp. 49–54.
- [5] K. Hanke, “Interframe wavelet video coding with lowpass transition,” in *Shanghai MPEG Meeting*, Oct. 2002, doc.m8997.
- [6] B. Pesquet-Popescu and V. Bottreau, “Three-dimensional lifting schemes for motion compensated video compression,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake City, UT, May 2001.
- [7] B. Girod and S. Han, “Optimum motion-compensated lifting,” *IEEE Signal Process. Lett.*, submitted for publication.
- [8] C. Tillier and B. Pesquet-Popescu, “3D, 3-band, 3-tap temporal lifting for scalable video coding,” in *Proc. IEEE Int. Conf. Image Process.*, Barcelona, Spain, Sep. 2003.
- [9] “3D MC-EZBC Software Package,” MPEG CVS Repository.